A nonparametric resampling for non causal random fields and its application to the texture synthesis

Lionel Truquet

Abstract

We study an extension to non causal Markov random fields of the resampling scheme given in Bickel et Levina (2006)[5] for texture synthesis with Markov mesh models. This extension is similar to a nonparametric method proposed by Paget and Longstaff (1998)[19] for texture synthesis and we also use their multiscale synthesis algorithm incorporating local annealing. We discuss some statistical properties and theoretical points for the convergence of the procedure and provide several convincing simulation examples.

1 Introduction

Over the two last decades, there was a particular attention to study the problem of texture synthesis. The goal of texture synthesis can be stated as follows: Given a texture sample, synthesize a new texture that appears similar to a human observer. Texture mapping or image compression are frequent applications for such algorithms. The stochastic nature of texture variations makes it a particularly natural area for applying statistical methods. The pioneer work of Cross and Jain (1983) [6] have shown the ability of Markov random fields to model a homogeneous texture. Such parametric models have been used for texture synthesis (as in [21]), but they require the estimation of a high number of parameters for capturing the complexity of real textures, which leads to computational difficulties. On the other hand, some algorithms model textures as a set of features, and generate new images by matching the features in an example texture ([7], [15], [23]). Those methods work very well for stochastic textures but have sometimes difficulties with highly structured ones.
Significant advances have been done in the area of texture synthesis using nonparametric algorithms with Markov random fields. A popular algorithm has been introduced by Efros and Leung (1999) [10]. Many variations of their method have been published that speed up and optimize the original algorithm in different ways in Wei et Levoy (2000) [24], Efros and Freeman (2001) [9] and Liang et al. (2001)[16] among others. The main statistical idea behind those algorithms is to consider the observed texture as a realization of a strictly stationary MRF. The data are used to construct an estimate of a local conditional distribution function of the field and a new texture is synthetized with a simulation procedure. Typically, the synthesis starts using a seed and pixels are synthetized in a given order by a recursive simulation of the random field as for time series. Intrinsically, those simulation procedures suppose the causality nature of the observed stochastic process. At a first sight this dependence form seems unnatural but the above algorithms work well on a wide variety of textures which seem well approximated by such random fields.
On the other hand, some noncausal procedures have been investigated. The FRAME model introduced by Zhu, Wu and Mumford [25] combine noncausal MRF models and feature matching. This last model has a mathematical justification: maximum entropy with empirical histograms of a finite number of filter responses are used to derive a parametric MRF for the whole distribution of the texture. Despite its solid statistical modeling, FRAME models does not work always very well on real textures.
Paget and Longstaff (1998)[19] have considered another algorithm, with a nonparametric noncausal

MRF. Contrarily to [25], the random field is specified through the conditional distribution and the empirical histogram is smoothed with a kernel which allows a simulation procedure with the Gibbs sampler. To avoid long relaxation time and phase discontinuities, Paget and Longstaff have used multiscale grids and have incorporated a temperature parameter for the pixels and the resulting algorithm is shown to be able to synthetize stochastic textures but also highly structured ones.
Except the work of Zhu et al.[25], not many theoretical works have been developed to study the consistency of such procedures. To our knowledge, the only contribution is the work of Bickel et Levina [5] who define a formal bootstrap scheme for resampling stationary (causal) random fields which gives a theoretical justification to the algorithm of Efros and Leung [10].
The goal of this paper is to extend the method of Bickel and Levina to noncausal random fields for modeling textures as in [19]. Of course, the use of the Gibbs sampler leads to long computational times and this gives a clear advantage to causal algorithms. However, from a theoretical point of view, the class of noncausal Markov random fields is known to be wider than the class of causal fields and in fact only a noncausal field has a real physical sense. In [5], the authors study a nonparametric estimation of the local conditional distribution function associated to the random field which is used to simulate an approximate causal field. This method is an extension to random fields of a $p-$order Markov bootstrap algorithm for time series [20]. We will use the same nonparametric estimation of a conditional law and we will use the multiscale synthesis algorithm given in [19] to give simulation examples.
The paper is organized as follows. In the following Section 2, we recall the results of Bickel et Levina and provide the natural extension of their method to the noncausal case. Some considerations on the convergence and the convergence rate of such algorithm are also provided. Section 3 is devoted to recall the multiscale algorithm used by Paget and Longstaff and we incorporate our bootstrap method to provide several simulation examples. Theoretical investigations are postponed to the two last sections of the paper.

2 The Markov Mesh Models algorithm

2.1 Principle

We first recall the Markov Mesh Models (MMM in sequel) algorithm introduced by Bickel and Levina [5]. This algorithm is different of the original algorithm of Efros and Leung [10] by the order in which pixels are filled in the synthesized texture (raster instead of spiral), and the weights with which the pixels are resampled. One can note that the raster order is used in some variations of the original algorithm (see [24] and [16]).

In all the sequel we consider $\{X_t,\, t \in \mathbb{N}^* \times \mathbb{N}^*\}$ a real-valued random field and a positive integer $o \in \mathbb{N}^*$. We will use the following notations:

- for $A \subset \mathbb{N}^* \times \mathbb{N}^*$, X_A denote the family $(X_t)_{t\in A}$;
- for $A,\, B \subset \mathbb{N}^* \times \mathbb{N}^*$, $A+B = \{t_A+t_B,\, (t_A,t_B) \in A\times B\}$ and $A-B = \{t_A-t_B,\, (t_A,t_B) \in A\times B\}$.

For $t = (t_1,t_2) \in \mathbb{N}^* \times \mathbb{N}^*$ and $s \in \mathbb{N}^* \times \mathbb{N}^*$, define the index sets

- $U_t^{(o)} = \{u = (u_1,u_2) \in \mathbb{N}^* \times \mathbb{N}^*;\ \max(1,t_1-o) \le u_1 \le t_1,\, \max(1,t_2-o) \le u_2 \le t_2 \text{ and } u \ne t\}$;
- $U_t^{(o)}(s) = U_t^{(o)} - \{t\} + \{s\}$;
- $W_t = \{1,\dots,t_1\} \times \{1,\dots,t_2\} \setminus \{t\}$.

The set $U_t^{(o)}$ is always included in the square of size $(o+1)\times(o+1)$ with t as the bottom right corner, t itself excluded, but there are $(o+1)^2-1$ possible shapes of $U_t^{(o)}$. Then,

Definition 1 *A random field* $X = \{X_t,\, t \in \mathbb{N}^* \times \mathbb{N}^*\}$ *is a Markov mesh model if there exists* $o \in \mathbb{N}^*$ *such that for all* $t \in \mathbb{N}^* \times \mathbb{N}^*$*,*

$$\mathbb{P}\left(X_t / X_{W_t}\right) = \mathbb{P}\left(X_t / X_{U_t^{(o)}}\right). \tag{2.1}$$

Now, the MMM resampling algorithm of Bickel and Levina [5] can be presented. First assume that a trajectory of a MMM X is observed on the index set $\{1, \ldots, T_1\} \times \{1 \ldots, T_2\}$ with $T_1, T_2 \in \{o, o+1, \ldots\} \times \{o, o+1, \ldots\}$, *i.e.*

$$(X_t,\, t \in \{1, \ldots, T_1\} \times \{1 \ldots, T_2\})$$

is known. Then consider a family of kernels $(W^{(\ell)})_{\ell \in \mathbb{N}^*}$ that are Borelian functions $W^{(\ell)} : \mathbb{R}^\ell \to [0, \infty)$ satisfying some general smoothness assumptions (see Assumption **(A4)** below). Moreover, for a resampling width $b > 0$ and all $\ell \in \mathbb{N}^*$, define

$$W_b^{(\ell)}(y) = b^{-\ell} W^{(\ell)}(y/b) \quad \text{for all } y \in \mathbb{R}^\ell.$$

In the sequel, for simplicity, we will omit the exponent o and ℓ for respectively $U_t^{(o)}$, $U_t^{(o)}(s)$ and $W_b^{(\ell)}$.

The MMM resampling algorithm

In the sequel we will denote $X^* = \{X_t^*,\, t \in \mathbb{N}^* \times \mathbb{N}^*\}$ the generated texture from $(X_t,\, t \in \{1, \ldots, T_1\} \times \{1 \ldots, T_2\})$. There are 3 main steps in this algorithm:

1. Select a starting value for $\{X_t^* : 1 \leq t_1 \leq o+1, 1 \leq t_2 \leq o+1\}$, the top left $(o+1) \times (o+1)$ square. Typically the starting value will be a $(o+1) \times (o+1)$ square random chosen from the observed field $(X_t,\, t \in \{1, \ldots, T_1\} \times \{1 \ldots, T_2\})$.

2. Suppose that there exists $(u, v) \in \mathbb{N}^* \times \mathbb{N}^*$ such that X_t^* has been generated for $t \in \{1, \ldots, u-1\} \times \{1, \ldots, v\} \cup \{u\} \times \{1, \ldots, v-1\}$, that is, $u-1$ rows are filled in completely, and the row u is filled up for the column v. To generate the next value $X_t^* = X_{(u,v)}^*$, let N_t be a discrete random variable with probability distribution

$$\mathbb{P}(N_t = s) = \frac{1}{Z} W_b\big(X_{U_t}^* - X_{U_t(s)}\big)$$

for all $s \in \mathbb{N}^* \times \mathbb{N}^*$ such that $U_t(s) \subset \{1, \ldots, T_1\} \times \{1 \ldots, T_2\}$ and where $Z = \sum_s W_b(Y_t^* - Y_t(s))$ is a normalizing constant. Note that the set of all possible s is such that all locations where the conditioning neighborhood fits within the observed texture field.

3. Generate N_t and set $X_t^* = X_{(u,v)}^* = X_{N_t}$.

In Figure 1, we show two steps in the progress of the MMM algorithm, just after the choice of the seed (step 1 above) and when the neighborhood of the pixel is full (see the center of the picture in Figure 1). In Figure 2, we illustrate the difference with the original algorithm of Efros and Leung. Here, the seed is put in the center of the new texture and the synthesis is done with a spiral ordering. To synthesize a pixel at a site t, one considers only the value of pixels already synthesized in a given square window centered at t. Another difference with the MMM algorithm is the choice of uniform weights for the synthesis (see [5] for details). The MMM algorithm is formulated for a particular class of random fields, the Markov Mesh Models (also known as Picard random fields) which were introduced by Abend, Harley and Kanal [1]. These models have been developed for image applications and can be simulated recursively an quickly. The resampling scheme described above is an adaption of a method proposed for bootstrapping Markovian time series ([22], [20]).

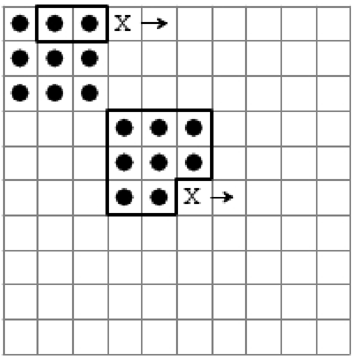


Figure 1: MMM algorithm, $o = 2$

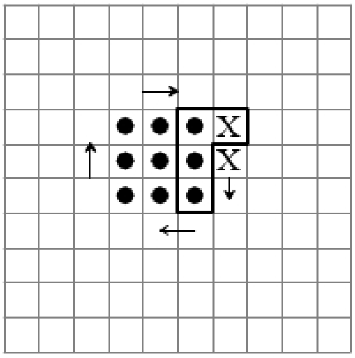


Figure 2: Efros and Leung algorithm, $o = 2$

2.2 Consistency results for causal models

First, let us introduce new notations:

- for $A \subset \mathbb{Z}^2$, $|A|$ is the cardinal of A.
- for $x = (x_1, x_2) \in \mathbb{Z}^2$, $\|x\|_\infty = \max(|x_1|, |x_2|)$.
- for $A, B \subset \mathbb{Z}^2$, $d(A, B) = \inf_{x \in A, y \in B}\{\|x - y\|_\infty\}$.
- for $y \in \mathbb{R}^\ell$ with $\ell \in \mathbb{N}^*$, $\|y\|$ is the usual Euclidian norm.
- for $T = (T_1, T_2) \in \mathbb{N}^* \times \mathbb{N}^*$, let $[T] = T_1 T_2$ and $T \to \infty$ means $T_1 \wedge T_2 \to \infty$.

Let $A \in \mathbb{Z}^2$ such that $|A| < \infty$. For $t \in \mathbb{Z}^2$, define

$$Y_t = (X_{t+j})_{j \in A}\,.$$

Moreover we define the following subsets of $\mathbb{N}^* \times \mathbb{N}^*$:

$$I_T = \Big\{t \in \{1, \ldots, T_1\} \times \{1 \ldots, T_2\},\ \{t\} - A \subset \{1, \ldots, T_1\} \times \{1 \ldots, T_2\}\Big\}.$$

To show the consistency of their algorithm, Bickel et Levina have proved a general lemma about the estimation of the local conditional distribution function

$$F_{X/Y}(x/y) = \mathbb{P}(X_t \le x / Y_t = y)$$

(see Theorem 2 in [5]). We first recall this theorem and its assumptions.

Assumptions of Theorem 2 in [5]

(A1) The random field X is strictly stationary and α-mixing, *i.e.* if for $k, u, v \in \mathbb{N}^*$,

$$\alpha_X(k, u, v) = \sup_{E, F \in \mathbb{Z}^2, d(E,F) = k, |E| = u, |F| = v} \Big\{|P(AB) - P(A)P(B)|,\ A \in \sigma(X_E), B \in \sigma(X_F)\Big\}$$

are the strong mixing coefficients such that there exist $\epsilon > 0$, $\tau > 2$ satisfying for all integers $u, v \geq 2$, $u + v \leq c$, where c is the smallest even integer such that $c \geq \tau$,

$$\sum_{k=1}^{\infty}(k+1)^{2(c-u+1)-1}\alpha_X(k,u,v)^{\epsilon/(c+\epsilon)} < \infty.$$

(A2) X_t has a compact support $S \subset \mathbb{R}$.

(A3) $F_{X,Y} = \mathbb{P}(X_t \leq \cdot, Y_t \leq \cdot)$, $F_{X/Y}$ and $F_Y = \mathbb{P}(Y_t \leq \cdot)$ have bounded continuous strictly positive densities (denoted $f_{X,Y}$, $f_{X/Y}$ and f_Y respectively) with respect to Lebesgue measure. Moreover, there exists $L > 0$ such that for any $y, y' \in S^A$, any $x \in S$,

$$\left| \int_{-\infty}^{x} f_{X,Y}(z,y)dz - \int_{-\infty}^{x} f_{X,Y}(z,y')dz \right| \leq L \left\| y - y' \right\|.$$

(A4) The family of kernels $(W^{(\ell)})_{\ell \in \mathbb{N}^*}$ is such that $W^{(\ell)} : \mathbb{R}^\ell \to (0, \infty)$ are bounded, symmetric and first-order Lipshitz continuous functions such that for all $\ell \in \mathbb{N}^*$,

$$\int u\, W^{(\ell)}(u)d\lambda_\ell(u) = 0 \quad \text{and} \quad \int \|u\|\, W^{(\ell)}(u)d\lambda_\ell(u) < \infty.$$

Moreover, the width of $W_b^{(\ell)}$ is supposed to be such that $b = b_T = O([T]^{-\delta})$, with $\delta > 0$.

To show the consistency of the MMM resampling algorithm, Bickel and Levina have established the convergence of the following sample cumulative conditional distribution function, that is, for $(x, y) \in S \times S^A$ and $T \in \mathbb{N}^* \times \mathbb{N}^*$ such that $I_T \neq \emptyset$:

$$F_T(x/y) = \frac{1}{Z_T} \sum_{s \in I_T} \mathbb{1}_{X_s \leq x}\, W_{b_T}(y - Y_s), \tag{2.2}$$

where $Z_T = \sum_{s \in I_T} W_{b_T}(y - Y_s)$.

Theorem 1 (Theorem 2 [5]) *If X is a MMM satisfying assumptions $(A1) - (A4)$, then for all $A \in \mathbb{Z}^2$ such that $|A| < \infty$,*

$$\sup_{(x,y) \in S \times S^A} \left| F_T(x/y) - F_{X/Y}(x/y) \right| \to_{T \to \infty} 0.$$

Theorem 1 shows the uniform convergence of the conditional distribution of a pixel given its neighborhood. Using this general result with neighborhoods A of causal nature (e.g. $A = U_t - \{t\}$), Bickel and Levina show the consistency of their MMM algorithm and also of the original spiral resampling algorithm of Efros and Leung. Their proof use the conditional independence properties of the MMM which allow a recursive computation of the joint laws (we refer to Theorem 1 in [5] for details). This resampling scheme uses the kernel regression estimation and requires some regularity assumptions (see Assumptions **(A1-4)**). As it is pointed in [5], those assumptions are perfectly plausible for most real textures: the mixing property is natural for stochastic textures, the compactness assumption is always satisfied since the number of gray levels is finite, and this number is sufficiently high in most of real textures to make the smoothness assumptions plausible.
However causal MMM are not really an appropriated for modeling texture: indeed, Why to choose a certain direction as for the dependence of the field? It is more natural to consider a spatial model for which there are no privileged direction for the dependence, *i.e.* a noncausal random field.

2.3 An extension to the noncausal case and a convergence rate of Theorem 1

To extend the previous results of [5] to noncausal fields, consider the following neighborhood $\mathcal{N}_o$ where $o \in \mathbb{N}^*$

$$\mathcal{N}_o = \{j \in \mathbb{Z}^2 / 0 < \|j\|_\infty \leq o\}.$$

Thus $\{t\} + \mathcal{N}_o$ is the natural extension of the set $U_t^{(o)}$ in the noncausal case. Denote

$$v = |\mathcal{N}_o| = (2o+1)^2 - 1.$$

The MMM is a very particular case of Markov random fields. If $X = \{X_t,\, t \in \mathbb{Z}^2\}$ is a $\mathbb{R}$-valued random field, then:

Definition 2 *$X = \{X_t,\, t \in \mathbb{Z}^2\}$ is a Markov random field if there exists $o \in \mathbb{N}^*$ such that for all $t \in \mathbb{Z}^2$,*

$$\mathbb{P}\left(X_t / X_{\mathbb{Z}^2 \setminus \{t\}}\right) = \mathbb{P}\left(X_t / X_{t+\mathcal{N}_o}\right). \tag{2.3}$$

We will again assume that $(X_t,\, t \in \{1, \ldots, T_1\} \times \{1 \ldots, T_2\})$ is known. Then, for all $t \in \mathbb{Z}^2$, define now:

$$Y_t = (X_{t+j})_{j \in \mathcal{N}_o} = X_{t+\mathcal{N}_o}.$$

First, a convergence rate for the Theorem 2 of [5] can be established and it is also satisfied in the noncausal case:

Theorem 2 *If X is a is a noncausal Markov random field satisfying assumptions $(\mathbf{A1-4})$, then for all $A \in \mathbb{Z}^2$ such that $|A| < \infty$,*

$$\sup_{(x,y) \in S \times S^{\mathcal{N}_o}} \left| F_T(x/y) - F_{X/Y}(x/y) \right| = O\left([T]^{-\gamma}\right) \qquad a.s$$

where $0 < \gamma < \dfrac{\tau - 2}{2(v+1)(\tau+v+2)}$ *and* $b = b_T = O([T]^{-\delta})$ *with* $\delta = \dfrac{\tau-2}{2(v+1)(\tau+v+2)}$.

Since MMM is a particular case of Markov random field, this result is also satisfied by MMM. It is interesting to see in both the causal or noncausal cases that the convergence rate of the MMM resampling algorithm is depending on a power law of $[T]$ (even if the choice of the optimal bandwidth b_T is depending on unknown parameters τ and v). Moreover the maximal exponent of convergence rate that we can obtain in Theorem 2 is $\frac{1}{2(1+v)}$ (that requires $\tau \to \infty$ for the mixing assumption). Remark that if $o = 0$ (corresponding to a independent random field) then $v = 0$ and the convergence rate is arbitrary close to $T^{1/2}$.

A partial consistency result for the MMM resampling algorithm in the noncausal case

In order to extend to the noncausal case the consistency proof of Bickel et Levina for the MMM resampling algorithm, we define the following one point conditional distribution defined by:

$$F_T(dx/y) = \frac{1}{Z_T} \sum_{s \in I_T} W_{b_T}\left(y - Y_s\right) \delta_{X_s}(dx), \tag{2.4}$$

where δ_x is the usual Dirac mass measure. Note that (2.4) is equal to (2.2) in the case $A = \mathcal{N}_o$. However and contrary to the causal case, the one point distribution (2.4) cannot be in general the one point conditional distribution of a noncausal Markov random field (nevertheless, we will use (2.4) to run a Gibbs sampler). A statistical problem with texture modeling by a noncausal Markov random field is to define a consistent nonparametric estimate of the one point conditional distributions which is also compatible with the existence of a conditional specification. This would allow to define an approximate Markov random field. We did not found a such estimate. Some tools are given in the Annex about the link between the convergence of a sequence of one point conditional distributions and the behavior of their joint laws provided they are well defined (see Theorem 3 and Theorem 4). Here we only provide a restrictive result of consistency of a simulation procedure directly with the Markov chain linked to the Gibbs sampler.

Suppose that we use the conditional distributions (2.4) and the Gibbs sampler to synthetize a new texture on a rectangle $R = R_T = \{1, \ldots, u_T\} \times \{1, \ldots, v_T\}$. We suppose here that assumptions of Theorem 1 hold. Though the conditional distributions F_T defined in (2.4) are not compatible with a Markov random field, we can use those distributions to simulate a Markov chain. We denote by $\prec$ the lexicographic order relation on R. Let z is an arbitrary element of $S^{\mathbb{Z}^2}$ not depending on T. If $x, y \in S^R$ and $s \in R$, we define the vectors $yx(s) \in S^{\mathcal{N}_o}$ such that $yx(s)_j = y_{s+j}$ if $s + j \prec s$ and $yx(s) = x_{s+j}$ otherwise, completed with the boundary conditions $x_{s+j} = z_{s+j}$ or $y_{s+j} = z_{s+j}$ if site $(s, j) \in R \times \mathcal{N}_o$ is such that $s + j \notin R$.

Now for $T \in \mathbb{N}^* \times \mathbb{N}^*$ such that $T_1, T_2 \geq 2o + 1$ (this ensures that I_T is not empty), we define the following transition on $X^R_{I_T} \subset S^R$:

$$P_T(x, dy) = \otimes_{s \in R} F_T\left(dy_s / yx(s)\right).$$

Note that P_T corresponds to the transition of the homogeneous Markov chain associated to the conditional distributions F_T when we implement the Gibbs sampler with a raster ordering for the visiting scheme (see [14] Theorem 6.2.1). Now as for the classical Gibbs sampler, we simulate a Markov chain on I^R_T, with initial value $w \in I^R_T$ and transition P_T. Since P_T is a positive transition, the law of this Markov chain with finite state space converges to its unique invariant probability denoted by μ_T. Then we have the following equality:

$$\mu_T(A) = \int P_T(x, A)\mu_T(dx), \qquad A \in \mathcal{B}(S^R),$$

where $\mathcal{B}(S^R)$ denote the Borel $\sigma-$algebra on S^R. One can mention that μ_T is not in general a measure that admits F_T as conditional distributions.
Since S^R is a compact metric space, the tightness of the sequence $(\mu_T)_T$ implies the existence of a cluster point denoted by μ. We are going to show that $\mu = \mu_R$, where μ_R denotes the conditional law $X_R / X_{\partial R} = z_{\partial R}$, where $\partial R = (R + \mathcal{N}_o) \setminus R$. Then by uniqueness of the cluster point, we will deduce the following consistency result:

$$\text{Almost surely : } \lim_{T \to \infty} \mu_T = \mu_R \quad \text{in distribution.} \tag{2.5}$$

To show that $\mu = \mu_R$, we first observe that μ_R is an invariant probability of the transition P on S^R defined by

$$P(x, dy) = \otimes_{s \in R} F_{X/Y}\left(dy_s / yx(s)\right), \quad x \in S^R.$$

Then P define a positive Markov chain and μ_R is the unique invariant probability. In fact P is the transition of the homogeneous Markov chain defined in the Gibbs sampler for the simulation of a realization of μ_R (still in the case of a periodic visiting scheme). Then if we prove that $\mu(A) = \int P(x, A)\mu(dx)$, $\forall A \in \mathcal{B}\left(S^R\right)$, we we can conclude that $\mu = \mu_R$.

Suppose that $(T_n)_{n \in \mathbb{N}}$ is sequence in $\mathbb{N}^* \times \mathbb{N}^*$ such that $\lim_{n \to \infty} \mu_{T_n} = \mu$. Then if g be a continuous and bounded function on S^R. We have:

$$\left| \int g(y)\mu_{T_n}(dy) - \int g(y)P(x, dy)\mu(dx) \right| \quad = \quad \left| \int g(y)P_{T_n}(x, dy)\mu_{T_n}(dx) - \int g(y)P(x, dy)\mu(dx) \right|$$

$$
\begin{aligned}
&\leq \sup_{x \in S^R} \left| \int g(y) P_{T_n}(x,dy) - \int g(y) P(x,dy) \right| \\
&+ \left| \int g(y) P(x,dy)(\mu_{T_n} - \mu)(dx) \right| \\
&= A + B.
\end{aligned}
$$

Since the function $x \to \int g(y)P(x,dy)$ is still bounded and continuous from assumption **(A2)**, the weak convergence of the sequence $(\mu_{T_n})_n$ implies that $B \to 0$ $(n \to \infty)$.
Now we show that $A \to 0$ $(n \to \infty)$. First we observe that if h is a continuous and bounded function on $S^R \times S^R$ and $s \in R$, then:

$$
\sup_{(x,y) \in S^R \times S^R} \left| \int h(x,y) \left(F_{T_n}(dy_s/yx(s)) - F_{X/Y}(dy_s/yx(s)) \right) \right| \to_{n \to \infty} 0, \quad \text{a.s.} \tag{2.6}
$$

The proof of (2.6) is omitted since the proof is very similar to the assertion $A \to 0$ in the proof of Theorem 3, using Theorem 1. If $u \in R$ is such that $u \succ s$, $\forall s \in R \setminus \{u\}$, we have:

$$
\begin{aligned}
&\left| \int g(y) P_{T_n}(x,dy) - \int g(y) P(x,dy) \right| \\
&\quad \leq \left| \int \int g(y) F_{X/Y}(dy_u/yx(u)) \left(\otimes_{s \in R \setminus \{u\}} F_{T_n}(dy_s/yx(s)) - \otimes_{s \in R \setminus \{u\}} F_{X/Y}(dy_s/yx(s)) \right) \right| \\
&\qquad + \sup_{y_s \in S, s \neq u} \left| \int g(y) \left(F_{T_n}(dy_u/yx(u)) - F_{X/Y}(dy_u/yx(u)) \right) \right|.
\end{aligned} \tag{2.7}
$$

Then if we iterate the bound (2.7), using a non increasing enumeration of the sites of R, the convergence $A \to 0$ follows from a repeated use of (2.6). Then, by the uniqueness of the limit of the sequence $(\mu_{T_n})_n$, we conclude that $\mu(dy) = \int P(x,dy)\mu(dx)$ and the convergence (2.5) follows from the previous remarks.

However, for obtaining the consistency the natural asymptotic requires that R_T increases to $\mathbb{Z}^2$. Unfortunately, we did not find a proof in this case.

3 The approach of Paget and Longstaff and simulation examples

3.1 Paget and Longstaff method

In their paper, Paget and Longstaff [19] have proposed a noncausal estimate of the local conditional distribution similar than ours, using also a kernel which smooths the multidimensional histogram. Our approach is more linked to the idea of a resampling scheme and appears in a natural way from the nonparametric estimation of the conditional expectation $(x,y) \mapsto \mathbb{P}\left(X_t \leq x / Y_t = y\right)$. The Gibbs sampler (see [12]) is a classical stochastic relaxation (SR) algorithm which is used for the simulation of Markov random fields. But as pointed in [19], a problem with the single-scale relaxation process is that global image characteristics evolve indirectly in the relaxation process. Global image characteristics are typically only propagated across the image lattice by local interactions and therefore evolve slowly, requiring long relaxation times to obtain equilibrium. Moreover the conditional distribution given in (2.4) requires the comparaison of a neighborhood in the output texture with all the neighborhoods of the same shape in the output texture. This leads to a very high computational load especially if p, the neighborhood size, must be very large to capture the global characteristics of the texture. This is why Paget and Longstaff used a multiscale relaxation, where the information obtained from SR at one resolution is used to constrain the SR at the next highest resolution. By this method, global image characteristics that have been resolved at a low resolution are infused into the

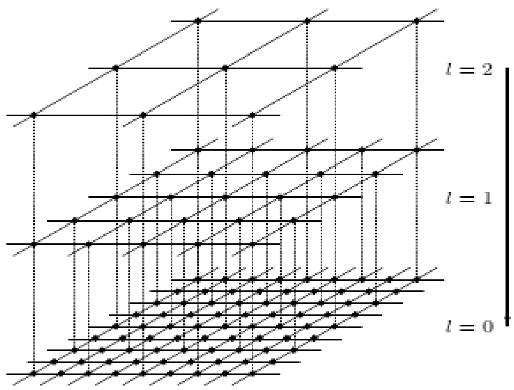


Figure 3: Grid organisation via decimation.

relaxation process at the higher resolutions. This helps to reduce the number of iterations required to obtain equilibrium with the Gibbs sampler.

The multigrid representation of an image is shown in Figure 3 which is taken from [19]. If $S_0 = [0, M_1] \times [0, M_2]$ represents the pixel's sites of an image x_0, the lower resolutions, or higher grid levels $l > 0$, are decimated versions of the image at level $l = 0$. For a grid level $l > 0$, the image x_l is defined on the lattice $S_l \subset S$, where

$$S_l = \{s = (2^l i, 2^l j)/0 \leq i \leq M_1/2^l, 0 \leq j \leq M_2/2^l\}.$$

The set of sites S_l at level l represents a decimation of the previous set of sites S_{l-1} at the lower grid level $l - 1$. The neighborhood system is redefined for each grid level $l > 0$:

$$\mathcal{N}_t^l = \{s \in S_l / \|t - s\|_\infty \leq o\}.$$

For level grid l, SR is not applied to the sites $s \in S^{l+1}$.

We refer to [17] for multiscale representations of Markov random fields. To better incorporate the multiscale relaxation described above, Paget and Longstaff have introduced a pixel temperature function used to determine when to terminate the SR process al one level and start it at the next level. Let l be a grid level. A pixel temperature is incorporated in equation (2.4) by modifying the form of the difference

$$d = y - Y_s. \tag{3.1}$$

In fact at the beginning of the SR at a level l, they define for a site $j \in S_l$ of the output texture the pixel temperature c_j as follows: $c_j = 0$ if $j \in S^{l+1}$ and $c_j = 1$ otherwise. The difference d is replaced by d' such that:

$$d'_j = (1 - c_{t+j})(x_{t+j} - X_{s+j}), \qquad j \in \mathcal{N}_o.$$

When a pixel x_t has been relaxed in the SR process, we set:

$$\widetilde{T}_t = \max\{0, \frac{\xi + \sum_{j \in \mathcal{N}_o} c_{t+j}}{|\mathcal{N}_o|}\}$$

where $\xi < 0$ is fixed by the user.

Here, the idea is to provide a total confidence to pixels coming from the preceding resolution and to progressively increase the confidence level of a pixel synthesized in the present resolution. When $c_j = 0 \ \forall j \in S_l$, the SR process is considered as having reached an equilibrium state indicating that the image can be propagated to the next lower grid level. This notion of temperature is related to the global temperature used in stochastic annealing (see [12]). Although we have incorporated this pixel temperature function for texture synthesis, we will not study in this paper statistical properties of a such approximation.

3.2 Texture synthesis examples

We have incorporated our noncausal bootstrap into the multiscale algorithm with the pixel temperature function described above. Concerning the choice of the parameters:

- For the neighborhood size, we choose $o = 3$ or $o = 4$.
- As in [5], we have not estimate the bandwidth parameter using theoretical results of kernel regression. We have empirically observed that $b = 0.01 \times (\text{neighborhood size})^{1/2}$ provides good results.
- As in [18], we set $\xi = -1$ and generally we have used 4 or 5 grid levels for the synthesis.

Another possibility for the simulation is to use a Conditional Iterative Mode (see [4]). The principle of this deterministic algorithm is to replace each step in the Gibbs sampler by choosing the value X_s such that $F_T\left(dx_t / x_{t+\mathcal{N}_o}\right)$ is maximal or equivalently such that $\|x_{t+\mathcal{N}_o} - Y_s\|$ is minimal in (2.4). Usually this algorithm converges toward a local extremum of the law of the random field on S^{R_T}. This local extremum depends on the initial values put for the pixels on the output texture. We have used the Conditional Iterative Mode for texture synthesis although its definition is not very clear in our case, since the joint laws are not defined.

To illustrate the principle of the multiscale algorithm, Figure 4 shows a step of the synthesis in the highest resolution. The Gibbs sampler runs in the raster ordering and Figure 4 shows the first sweep. One can see that the lower resolutions give the shape of the texture. Moreover the pixel temperature function helps for a good initialization of the sampler.
This multiscale algorithm does not correctly work only for stochastic textures as in Figure 8 but also for highly structured ones as Figure 5 shows, even if small discontinuities appear in the last case.
In fact, we have observed that the ICM works as well as the non deterministic algorithm and in some cases better as in Figure 6.
Figure 9 exhibits a comparison with Efros and Leung's algorithm in a failure case. Texture (b) is taken from Efros and Leung's paper [10] and shows that this causal approach can create garbage when the algorithm slips into a wrong part of the search space. Although the noncausal algorithm does not have the same problem, a gray dark area is often reproduced in texture (c).
Figure 10 shows a comparison with two some populars pixel by pixel algorithms. Texture (b) synthesized using Wei et Levoy algorithm [24] and texture (c) using the Ashikhmin method are taken from [2]. The noncausal algorithm used for texture (d) avoids excessive blurring as in (b) and rough images as in (c).

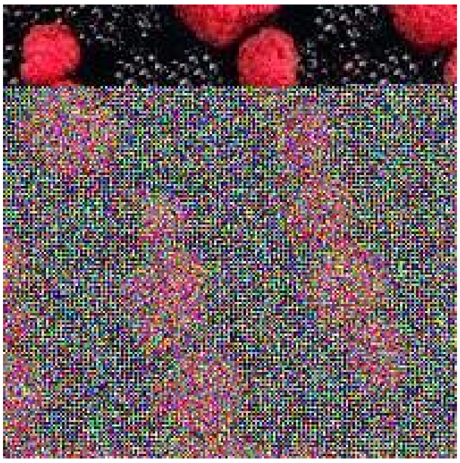
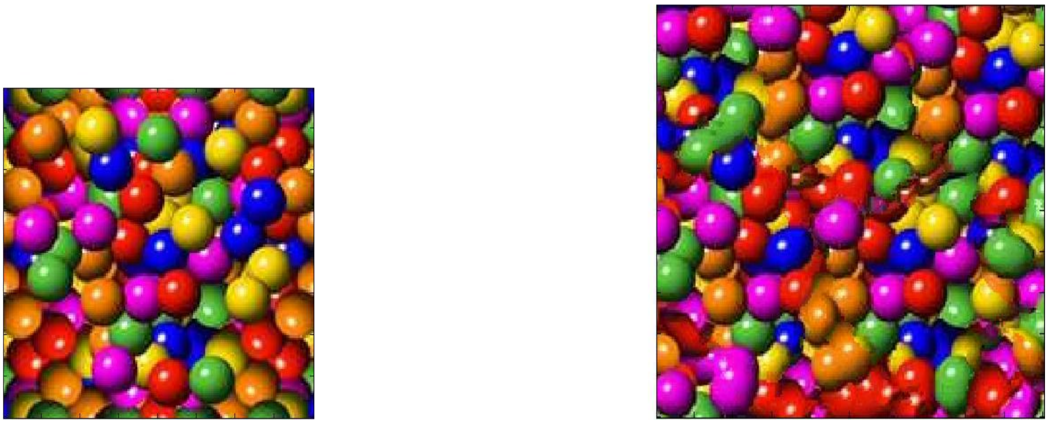
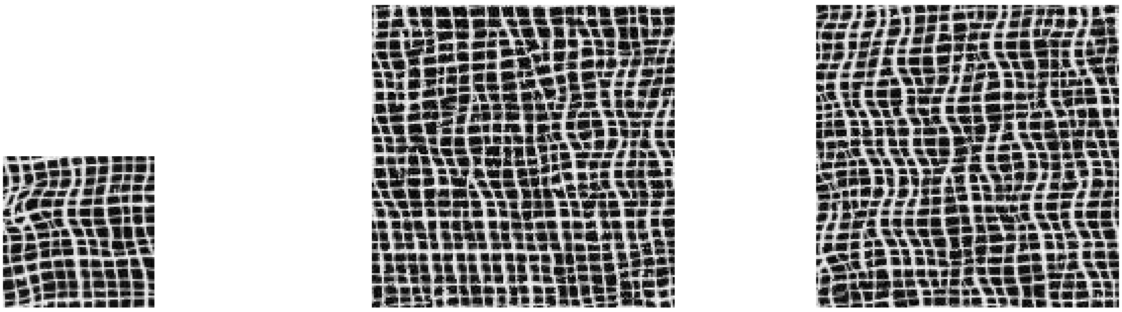


Figure 4: The Gibbs sampler and the highest resolution.



(a) (b)

Figure 5: (a) Original texture 160×160 pixels, (b) Synthesis with the multiresolution algorithm 200×200 pixels.



(a) (b) (c)

Figure 6: (a) sample 75×75 pixels, (b) multiresolution algorithm 150×150, (c) ICM 150×150.

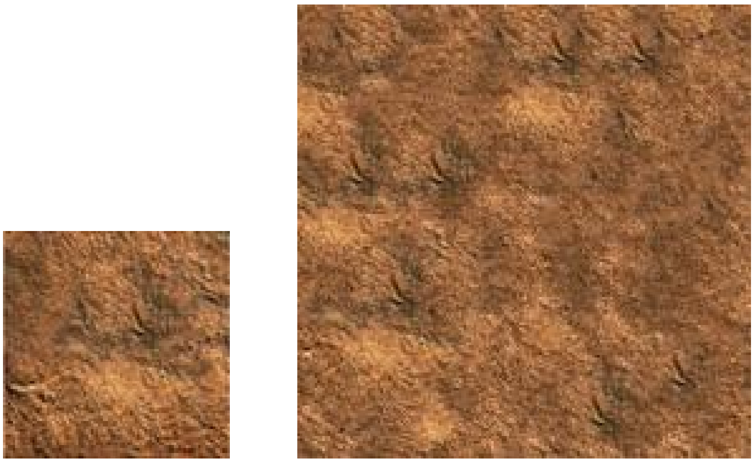


Figure 7: Original texture (128 × 128) and synthesis (200 × 200)

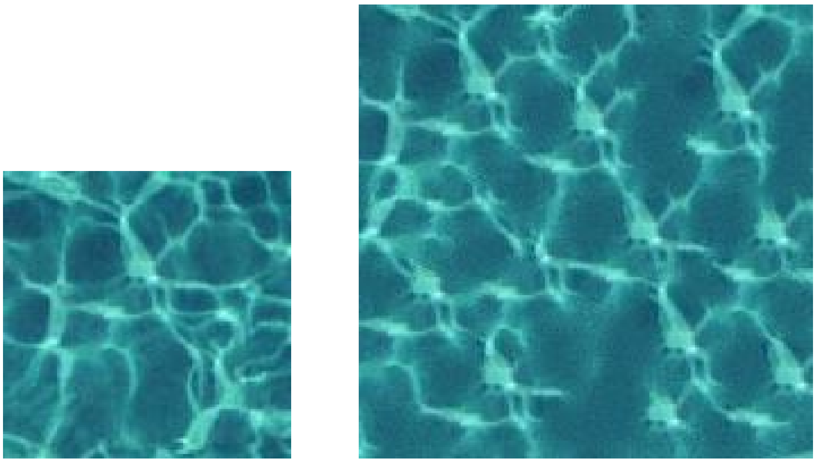


Figure 8: Original texture (128 × 128) and synthesis (200 × 200)

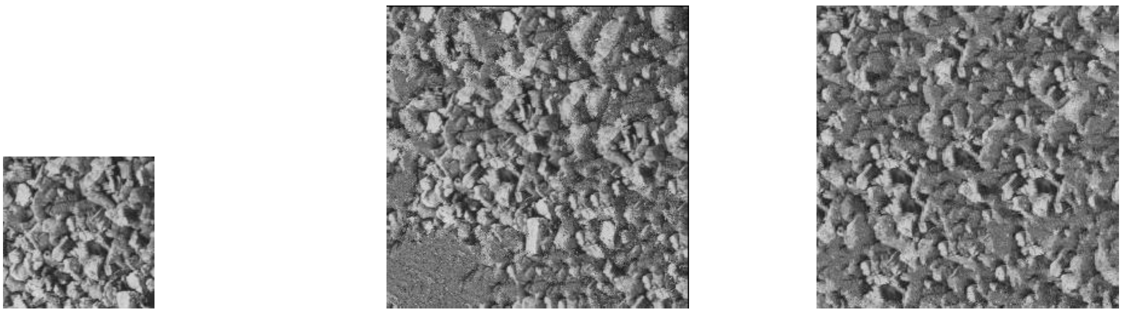
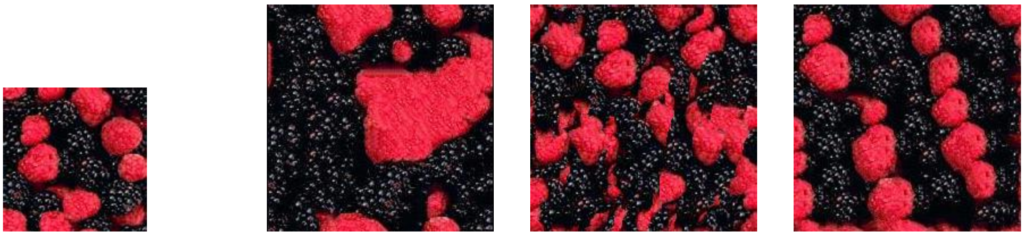


Figure 9: (a) Sample 128 × 128 pixels, (b) Efros and Leung's result, (c) Our method with $o = 3$ (250 × 250 pixels in each case).



(a) (b) (c) (d)

Figure 10: Comparison with causal methods: (a) sample 128×128, (b) Wei et Levoy algorithm, (c) Ashikhmin method, (d) ICM. All the synthesized textures are 200×200 pixels.

4 Annex

We will use the convenient notation: for $s,t\in\mathbb{N}^*\times\mathbb{N}^*$ define:

$$Y_t = X_{U_t^{(o)}} \quad\text{and}\quad Y_t(s) = X_{U_t^{(o)}(s)}.$$

4.0.1 Proof of Theorem 2

We follow the proof of theorem 2 of Bickel et Levina in order to compute convergence rate. We first recall the following lemma which proof can be found in [8].

Lemma 1 *(Moment inequality). Let F_t be a real-valued random field indexed by $I\subset\mathbb{Z}^d$ satisfying conditions* **(A1)**. *If $\mathbb{E}F_t=0$, $F_t\in\mathbb{L}^{\tau+\epsilon}$ and $\tau\geq 2$, then there exists a constant C depending only on τ and mixing coefficients of F_t such that*

$$\mathbb{E}\left|\sum_{t\in I}F_t\right|^{\tau} \leq C\max\left(L(\tau,\epsilon), L(2,\epsilon)^{\tau/2}\right),$$

where

$$L(\mu,\epsilon)=\sum_{t\in I}\left(\mathbb{E}\left|F_t\right|^{\mu+\epsilon}\right)^{\mu/(\mu+\epsilon)}.$$

It is easy to see that if $\sup_t\|F_t\|_\infty\leq M$, then we obtain:

$$\mathbb{E}\left|\sum_{t\in I}F_t\right|^{\tau}\leq CM^{\tau}\left|I\right|^{\tau/2} \tag{4.1}$$

For $(x,y)\in S\times S^A$, we set:

$$r_T(x,y)=[T]^{-1}\sum_{s\in I_T}\mathbb{1}_{(-\infty,x]}(X_s)W_b(y-Y_s),\quad r(x,y)=\int\mathbb{1}_{(-\infty,x]}(z)f_{X,Y}(z,y)dz,$$

$$f_T(y)=[T]^{-1}\sum_{s\in I_T}W_b(y-Y_s).$$

We have:

$$F_T(x/y)=\frac{r_T(x,y)}{f_T(y)},\quad F_{X/Y}(x/y)=\frac{r(x,y)}{f_Y(y)}. \tag{4.2}$$

Following the proof of lemma A2 in [5], we prove the following result

Lemma 2 *Under assumptions $(A1)-(A4)$, for any $x\in\mathbb{R}$*

$$\sup_{(x,y)\in S\times S^A}\left|r_T(x,y)-r(x,y)\right| = O([T]^{-\gamma})$$

for $0<\gamma<\frac{\tau-2}{2(v+1)(\tau+v+2)}$.

Proof of Lemma 2 In this proof, we will denote by $C > 0$ a generic constant which does not depend on T.

Let $\delta > 0$ such that $b_T = O\left([T]^{-\delta}\right)$, then the proof of lemma A2 in [5] leads to

$$\sup_{(x,y)\in S\times S^A} |\mathbb{E}r_T(x,y) - r(x,y)| = O\left([T]^{-\delta}\right). \tag{4.3}$$

Then we need to bound $\sup_{(x,y)\in S\times S^A} |r_T(x,y) - \mathbb{E}r_T(x,y)|$.
As in [5], we define

$$Z_{t,T}(x,y) = \mathbb{1}_{(-\infty,x]}(X_t)W_{b_T}(y - Y_t) - \mathbb{E}\left(\mathbb{1}_{X_t\le x}W_{b_T}(y - Y_t)\right)$$

and we need to bound $\sup_{(x,y)\in S\times S^A}\left|\frac{1}{[T]}\sum_{t\in I_T} Z_{t,T}(x,y)\right|$.
As $S\times S^A$ is compact, we can cover $S\times S^A$ with N_T cubes $I_{i,T}$ with centers (x_i, y_i) and sides L_T for the supremum norm. Without loss of generality, we suppose $x_1 \le \dots \le x_{N_T}$ and we set $x_0 = x_1 - L_T$ and $x_{N_T} = x_{N_T} + L_T$. Then

$$\begin{aligned}
\sup_{(x,y)\in S\times S^A}\left|[T]^{-1}\sum_{t\in I_T} Z_{t,T}(x,y)\right| &\le \max_{1\le i\le N_T}\left|[T]^{-1}\sum_{t\in I_T} Z_{t,T}(x_i,y_i)\right| \\
&+ \max_{1\le i\le N_T}\sup_{(x,y)\in(S\times S^A)\cap I_{i,T}}\left|[T]^{-1}\sum_{t\in I_T}\left(Z_{t,T}(x,y) - Z_{t,T}(x_i,y_i)\right)\right| \\
&= I + II
\end{aligned}$$

- First let deal with term II. Using assumption **(A4)** for the kernel, we have for $t \in I_T$ and $x \in (x_{i-1}, x_i]$:

$$\begin{aligned}
|Z_{t,T}(x,y) - Z_{t,T}(x_i,y_i)| &\le C\left(b_T^{-(v+1)}\|y - y_i\| + b_T^{-v}\left(\mathbb{1}_{x_{i-1}<X_t\le x_i} + \mathbb{P}(x_{i-1} < X_t \le x_i)\right)\right) \\
&\le C\left(b_T^{-(v+1)}L_T + b_T^{-v}\left(\mathbb{1}_{x_{i-1}<X_t\le x_i} + \mathbb{P}(x_{i-1} < X_t \le x_i)\right)\right)
\end{aligned}$$

  We choose $L_T = [T]^{-\beta}$ and we set $U_{i,t} = \mathbb{1}_{]x_{i-1},x_i]}(X_t) - \mathbb{P}(x_{i-1} < X_t \le x_{i+1})$. Remark that assumption **(A2)** about the existence of densities allows to derive the bound:

$$\mathbb{P}(x_{i-1} < X_0 \le x_i) \le CL_T.$$

  We have:

$$\begin{aligned}
&\sup_{(x,y)\in(S\times S^A)\cap I_{i,T}}\left|[T]^{-1}\sum_{t\in I_T}\left(Z_{t,T}(x,y) - Z_{t,T}(x_i,y_i)\right)\right| \\
&\le C\left([T]^{\delta(v+1)-\beta} + [T]^{v\delta-1}|\sum_{t\in I_T} U_{i,t}| + [T]^{v\delta}\mathbb{P}(x_{i-1} < X_0 \le x_i)\right) \\
&\le C\left([T]^{\delta(v+1)-\beta} + \max_{1\le i\le N_T}[T]^{v\delta-1}|\sum_{t\in I_T} U_{i,t}|\right).
\end{aligned}$$

  Now we consider a real number $\gamma < \frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}$. Since $N_T = O\left(L_T^{-(v+1)}\right) = O\left([T]^{\beta(v+1)}\right)$, we obtain using (A1) and lemma 1

$$\mathbb{P}\left(\max_{1\le i\le N_T}[T]^{v\delta-1}\left|\sum_{t\in I_T} U_{i,t}\right| > [T]^{-\gamma}\right) \le \sum_{i=1}^{N_T}[T]^{(\gamma+v\delta-1)\tau}\mathbb{E}\left|\sum_{t\in I_T} U_{i,t}\right|^{\tau}$$

$$\leq \quad C[T]^{\beta(v+1)+(\gamma+v\delta-1)\tau+\frac{\tau}{2}}$$

By the choice of γ, we have $\beta(v+1)+(\gamma+v\delta-1)\tau+\frac{\tau}{2}<-1$ and we deduce from the Borel Cantelli lemma that

$$\max_{1\leq i\leq N_T}[T]^{v\delta-1}\left|\sum_{t\in I_T}U_{i,t}\right|=O\left([T]^{-\gamma}\right)\text{ a.s,}\quad \gamma<\frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}.$$

Now using the previous inequalities, we deduce that:

$$II\leq O\left([T]^{\delta(v+1)-\beta}+[T]^{-\gamma}\right),\quad \gamma<\frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}. \tag{4.4}$$

- Now we turn on the term I. For a real number $\widetilde{\gamma}<\frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}$, we have using (A1) and lemma 1:

$$\begin{aligned}\mathbb{P}(I>[T]^{-\widetilde{\gamma}}) &\leq \sum_{i=1}^{N_T}[T]^{(\widetilde{\gamma}-1)\tau}\mathbb{E}\left|\sum_{t}Z_{t,T}(x_i,y_i)\right|^{\tau}\\ &\leq CN_T[T]^{(\widetilde{\gamma}-1)\tau+\frac{\tau}{2}}b_T^{-v\tau}\\ &\leq C[T]^{\beta(v+1)+(\widetilde{\gamma}-1)\tau+\frac{\tau}{2}+\delta v\tau}\end{aligned}$$

By the choice of $\widetilde{\gamma}$, we have

$$\beta(v+1)+(\widetilde{\gamma}-1)\tau+\frac{\tau}{2}+\delta v\tau<-1,$$

and by the Borel Cantelli lemma, we have

$$I=O\left(T^{-\widetilde{\gamma}}\right)\quad\text{a.s,}\quad \widetilde{\gamma}<\frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}$$

Now we choose the number β such that:

$$\beta-\delta(v+1)=\frac{\tau-2-2v\delta\tau-2\beta(v+1)}{2\tau}.$$

This leads to $\beta=\frac{\tau-2+2\tau\delta}{2(v+\tau+1)}$ and to the following rate:

$$I+II=O\left([T]^{-\gamma}\right),\quad \gamma<\frac{\tau-2-2\delta((v+1)^2+v\tau)}{2(v+\tau+1)}.$$

Finally, we choose δ for an equilibrium with the bound (4.3), solving the equation:

$$\delta=\frac{\tau-2-2\delta((v+1)^2+v\tau)}{2(v+\tau+1)}.$$

This leads to:

$$\delta=\frac{\tau-2}{2(v+1)(\tau+v+2)}>0,$$

which gives the rate given by the Lemma 2.

□

Proof of Theorem 2 We write:

$$\begin{aligned}\left|F_T(x/y) - F_{X/Y}(x/y)\right| &= \frac{1}{f_T(y)}\left|r_T(x,y) - r(x,y) + r(x,y) - F_{X/Y}(x/y)f_T(y)\right| \\ &\leq \frac{1}{f_T(y)}\left(|r_T(y) - r(y)| + F_{X/Y}(x/y)\left|f_T(y) - f_Y(y)\right|\right)\end{aligned}$$

By lemma 2, we have $\sup_{x,y}|r_T(x,y) - r(x,y)|, \sup_y|f_T(y) - f(y)| = O\left(T^{-\gamma}\right)$ a.s with $0 < \gamma < \frac{\tau-2}{2(v+1)(\tau+v+2)}$. Since by (A2), $\inf_{y\in S^A} f_Y(y) > 0$ and $\sup_{(x,y)\in S\times S^A} F_{X/Y}(x/y) \leq 1$, we get the result.□

5 Some tools for the consistency: Continuity results

In this section, we give two results which describe the behavior of random fields in relation to their one point conditional distributions. If it is possible to construct a nonparametric estimate of the one point conditional distribution which is also compatible with the one point conditional distribution of a Markov random field, the following results will be useful to describe the statistical properties of the joint laws of the model.

We first give a lemma which states the behavior of the joint laws of a sequence of random fields when their one-point conditional distributions are convergent. In the sequel, let S be a compact set of $\mathbb{R}$ endowed with its Borelian algebra $\mathcal{B}(S)$. Let $\mathcal{X} = S^I$ where I is a denumerable set. For any sequence $(u_i)_{i\in I}$ of positive real numbers satisfying $\sum_{i\in I} u_i < \infty$, we consider the distance d on $\mathcal{X}$ defined by:

$$d(z,z') = \sum_{i\in I} u_i\left|z_i - z'_i\right|, \qquad z, z' \in \mathcal{X}.$$

Then $(\mathcal{X}, d)$ is a compact metric space. For $A \subset I$, let $p_A : \mathcal{X} \to S, \quad z \mapsto z_A$. For $t \in I$, we will write p_t instead of $p_{\{t\}}$. Moreover, for $t \in I$, we set:

$$\mathcal{F}_t^- = \sigma(p_j/ \quad j \neq t).$$

We denote by $\mathcal{P}(\mathcal{X})$ the set of probability measures on $\mathcal{X}$. If ν_1, ν_2 are two elements of $\mathcal{P}(\mathcal{X})$, the Prohorov distance d_P between ν_1 and ν_2 is defined by:

$$d_P(\nu_1,\nu_2) = \inf\{\epsilon > 0, \nu_1(A) \leq \nu_2(A^\epsilon) + \epsilon, \forall A \in \mathcal{B}(\mathcal{X})\},$$

where $A^\epsilon = \{z \in \mathcal{X}/d(z,A) \leq \epsilon\}$. The distance d_P defines the weak convergence on $\mathcal{P}(\mathcal{X})$ which is a compact space topology.
Now for $\nu \in \mathcal{P}(\mathcal{X})$ and any bounded measurable function f on $\mathcal{X}$, we set:

$$\mathbb{E}_\nu(f) = \int f d\nu.$$

For $t \in I$, we denote ν_t the kernel on $\mathcal{P}(\mathcal{X})$ such that:

$$\nu_t(A/z) = \mathbb{E}_\nu\left(\mathbb{1}_A/\mathcal{F}_t^-\right)(z),$$

where $\mathbb{E}_\nu\left(\cdot/\mathcal{F}\right)$ denotes the conditional expectation with respect to a $\sigma-$algebra $\mathcal{F} \subset \mathcal{B}(\mathcal{X})$.

Finally let $\gamma = (\gamma_t)_{t\in I}$ be a sequence of probability kernels such that for $t \in I$, γ_t is a kernel from $\mathcal{F}_t^-$ to $\mathcal{B}(\mathcal{X})$ satisfying the property:

$$\gamma_t\left(A \cap B/\cdot\right) = \gamma_t\left(A/\cdot\right) \times \mathbb{1}_B, \quad (A,B) \in \sigma(p_t) \times \mathcal{F}_t^-.$$

If h is a bounded measurable function on $\mathcal{X}$, we denote $\gamma_t(h)$ the measurable function on $\mathcal{X}$ such that:

$$\gamma_t(h)(z) = \int f(w)\gamma_t(dw/z), \qquad z \in \mathcal{X}.$$

We define the following subset of $\mathcal{P}(\mathcal{X})$:

$$\mathcal{G}(\gamma) = \{\nu \in \mathcal{P}(\mathcal{X}) / \quad \forall t \in I, \quad \nu\gamma_t = \nu\},$$

where for all $(t,\nu) \in I \times \mathcal{P}(\mathcal{X})$, $\nu\gamma_t$ denotes the element of $\mathcal{P}(\mathcal{X})$ such that:

$$\nu\gamma_t(A) = \int \gamma_t(A/z)d\nu(z).$$

We will say that γ satisfies the condition **(C)** if:

$$(\mathbf{C}) \qquad \forall t \in I, \quad h \in \mathcal{C}(\mathcal{X}) \Rightarrow \gamma_t(h) \in \mathcal{C}(\mathcal{X}),$$

where $\mathcal{C}(\mathcal{X})$ is the space of continuous and bounded functions on $\mathcal{X}$.
The following result gives the behavior of a sequence of random fields in the case of uniform convergence of their one point conditional distribution. A general treatment of topological properties of random fields is given in [13]. For completeness of this work, we state and prove the following result:

Theorem 3 *For $t \in I$, let γ_t be a probability kernel on $\mathcal{X} \times \mathcal{F}_t^-$. Suppose that the sequence $\gamma = (\gamma_t)_{t\in I}$ satisfies condition* **(C)***. Then for a sequence $(\nu^{(n)})_n$ of $\mathcal{P}(\mathcal{X})$ such that*

$$\sup_{(x,z)\in S\times\mathcal{X}} \left|\nu_t^{(n)}(p_t \leq x/z) - \gamma_t(p_t \leq x/z)\right| \to_{n\to\infty} 0,$$

we have $d_P(\nu^{(n)}, \mathcal{G}(\gamma)) \to_{n\to\infty} 0$.

Proof of Lemma 3 Suppose that there exists $\epsilon > 0$ and a subsequence $s = (\nu^{(n_k)})_{k\in\mathbb{N}}$ such that $d_P(\nu_{n_k}, \mathcal{G}(\gamma)) > \epsilon$, $\forall k \in \mathbb{N}$. Since this sequence is relatively compact with respect to the weak topology, then there exists a subsequence $(\nu^{(n_{k'})})_{k'}$ of s and $\nu \in \mathcal{P}(\mathcal{X})$ such that:

$$\lim_{k'\to\infty} \nu^{(n_{k'})} = \nu.$$

We are going to show that $\nu \in \mathcal{G}(\gamma)$, which is a contradiction with $d_P(\nu^{n_{k'}}, \mathcal{G}(\gamma)) > \epsilon$, $k' \in \mathbb{N}$.
Let $h \in C(\mathcal{X})$. For $t \in I$, we have:

$$\begin{aligned} & \int h d\nu^{(n_{k'})} - \int h d(\nu\gamma_t) \\ = & \int \left(\nu_t^{(n_{k'})}(h) - \gamma_t(h)\right) d\nu^{(n_{k'})} + \int \gamma_t(h) d\nu^{(n_{k'})} - \int \gamma_t(h) d\nu \\ = & A + B. \end{aligned}$$

Let $\varepsilon > 0$. Since h is uniformly continuous on $\mathcal{X}$, there exists $\delta > 0$ such that $d(z,z') < \delta \Rightarrow |h(z) - h(z')| < \epsilon$. We choose a subdivision $a_1, \ldots, a_k$ of the interval $[a,b] \supset S$ with step smaller then δ/u_t. Let h_k the function defined on $[a,b]$ by $h_k(z) = \sum_{l=1}^{k-1} h(z(l))\mathbb{1}_{]a_l,a_{l+1}]}(z_0)$ where for $l = 1,\ldots,k-1$, $z(l)$ is the element of $\mathcal{X}$ such that $z(l)_t = a_l$ and $z(l)_s = z_s$ if $s \neq t$. We have $\sup_{z\in\mathcal{X}} |h(z) - h_k(z)| < \epsilon$.
We deduce:

$$\left|\nu_t^{(n_{k'})}(h)(z) - \gamma_t(h)(z)\right| \leq 2\epsilon + 2(k-1)\|h\|_\infty \sup_{(x,z)\in S\times\mathcal{X}} \left|\nu_t^{(n)}(p_t \leq x/z) - \gamma_t(p_t \leq x/z)\right|.$$

One can conclude that:

$$\sup_{z\in\mathcal{X}}\left|\nu_t^{(n_{k'})}(h)(z)-\gamma_t(h)(z)\right|\rightarrow_{k'\to\infty}0.$$

Hence $A\to 0$.

Moreover since $\gamma_t(h)\in C(\mathcal{X})$ by condition **(C)**, the weak convergence of the sequence $\left(\nu^{(n_{k'})}\right)_{k'\in\mathbb{N}}$ implies $B\to 0$.

Then we conclude that for $t\in I$, we have $\nu\gamma_t=\nu$, and the result follows from this contradiction.□

Now we investigate the following problem. Suppose for simplicity that $S=[a,b]$. Assume that for $\mu,\nu\in\mathcal{P}(\mathcal{X})$, the distance between the conditional distribution functions $\nu_t(p_t\leq\cdot/\cdot)$ and $\mu_t(p_t\leq\cdot/\cdot)$ is known, then is it possible to obtain the distance between μ and ν over some cylinders sets of the form

$$C_{x_{t_1},\dots,x_{t_k}}=\otimes_{i=1}^k[a,x_{t_i}]\times S^{I\backslash\{t_1,\dots,t_k\}}?$$

In other words, can we obtain the distance between the distribution functions of the joint laws? This problem is linked to the phase transition phenomenon and to the Dobrushin's contraction formula. In order to apply this formula, the following assumption will be needed:

(H) We assume that there exist two families of non negative real numbers $\{L_{t,j}/t,j\in I\}$ and $\{M_{t,j}/t,j\in I\}$, with $L=\sup_{t\in I}\sum_{j\neq t}L_{t,j}<1$ and $M=\sup_{t\in I}\sum_{j\neq t}M_{t,j}<\infty$, such that $\forall z,z'\in\mathcal{X}$:

$$\sup_{x\in S}\left|\mu\left(p_t\leq x/z\right)-\mu\left(p_t\leq x/z'\right)\right|\leq\sum_{j\neq t}M_{t,j}\left|z_j-z'_j\right|,$$

$$\int_S\left|\mu\left(p_t\leq x/z\right)-\mu\left(p_t\leq x/z'\right)\right|dx\leq\sum_{j\neq t}L_{t,j}\left|z_j-z'_j\right|.$$

Theorem 4 *Let $\mu,\nu\in\mathcal{P}(\mathcal{X})$ with $S=[a,b]$. Suppose that the random field μ satisfies assumption* **(H)**. *Then for each finite subset $\{t_1,\dots,t_k\}$ of I, we have:*

$$\sup_{(x_{t_1},\dots,x_{t_k})\in S^k}\left|\mu\left(C_{x_{t_1},\dots,x_{t_k}}\right)-\nu\left(C_{x_{t_1},\dots,x_{t_k}}\right)\right|\leq C\sup_{t\in I}\sup_{(x,z)\in S\times\mathcal{X}}\left|\mu(p_t\leq x/z)-\nu(p_t\leq x/z)\right|,$$

where $C=1+M(b-a)\left(k-1+\frac{1}{1-L}\right)$.

We first recall the following inequality due to Dobrushin (see [11] remark 2.17). This inequality allows to bound the distance between two random fields μ and ν with the distance between their local conditional specification. Of course, a such inequality implies that there is no phase transition. Some contraction conditions on the local conditional specifications are needed to get this inequality. Before giving this inequality in our case, we introduce some notations. Let r be a metric on $S=[a,b]$. If α and β are two probability on S, the Warsserstein metric is defined as

$$R(\alpha,\beta)=\sup\frac{\left|\int fd\alpha-\int fd\beta\right|}{\delta(f)},$$

where the supremum is taken over all Lipshitz functions f on S with

$$\delta(f)=\sup_{x\neq x'}\frac{|f(x)-f(x')|}{r(x,x')}<\infty.$$

For our result, we will only consider the metric $r(x,x')=|x-x'|$. One can mention that in this case, R has the following expression:

$$R(\alpha,\beta)=\int|\alpha([a,x])-\beta([a,x])|\,dx,\quad \alpha,\beta\in\mathcal{P}(S). \tag{5.1}$$

We define also $L(\mathcal{X})$ the space of real functions f such that:

$$\left|f(z)-f(z')\right|\le\sum_{i\in\mathbb{Z}^2}\left|z_i-z_i'\right|\delta_i(f),\quad \sum_{i\in\mathbb{Z}^2}\delta_i(f)<\infty$$

where

$$\delta_i(f)=\sup_{z\neq z'}\left\{\frac{|f(z)-f(z')|}{|z_i-z_i'|}/z_j=z_j',\forall j\neq i\right\}.$$

Let $\mu,\nu\in\mathcal{P}(\mathcal{X})$. We suppose that the following continuity condition holds:

$$f\in L(\mathcal{X})\Rightarrow\forall t\in\mathbb{Z}^2,\mu_t(f)\in L(\mathcal{X}). \tag{5.2}$$

For $\mu\in\mathcal{P}(\mathcal{X})$, the contraction coefficients are defined by

$$C_{ik}=\sup\left\{\frac{R\left(\mu(p_k\in\cdot/z),\mu(p_k\in\cdot/z')\right)}{|z_i-z_i'|}/z_j=z_j',\forall j\neq i\right\}$$

and

$$b_k=\int R\left(\mu(p_k\in\cdot/z),\nu(p_k\in\cdot/z)\right)\nu(dz)$$

Note that with the expression (5.1), we have the bound:

$$b_k\le(b-a)\sup_{(x,z)\in S\times\mathcal{X}}\left|\mu(p_k\le x/z)-\nu(p_k\le x/z)\right|. \tag{5.3}$$

Let $D=\sum_{n\ge0}C^n$ where C^n denotes the nth power of the matrix C. D is well defined for example if

$$c=\sup_{k\in\mathbb{Z}^2}\sum_{i\in\mathbb{Z}^2}C_{ik}<1. \tag{5.4}$$

In this case, the following inequality holds:

$$\left|\int fd\mu-\int fd\nu\right|\le\sum_{i\in\mathbb{Z}^2}(bD)_i\delta_i(f),\quad f\in L(\mathcal{X}). \tag{5.5}$$

Proof of Theorem 4 First, from the point 1 of lemma 3, condition (5.2) is satisfied for μ. Moreover for $i,k\in\mathbb{Z}^2$, we have $C_{i,k}=L_{k,i}$ and the condition (5.4) is satisfied with $c=L$. Then, inequality (5.5) holds for ν.

For $l\in\{1,\dots,k\}$ and $x\in\mathcal{X}$, we set:

$$f_l(z)=\prod_{m=1}^{l}\mathbb{1}_{(-\infty,x_{t_m}]}(z_{t_m})$$

and

$$g_l=\mu_{t_l}\circ\cdots\circ\mu_{t_1}(f_l),\quad h_l=\nu_{t_l}\circ\cdots\circ\nu_{t_1}(f_l).$$

We have:

$$\left|\int f_kd\mu-\int f_kd\nu\right|\;=\;\left|\int g_kd\mu-\int h_kd\nu\right|$$

$$\begin{aligned} &\leq \left|\int g_k d\mu - \int g_k d\nu\right| + \left|\int g_k d\nu - \int h_k d\nu\right| \\ &\leq \left|\int g_k d\mu - \int g_k d\nu\right| + \beta_k \end{aligned}$$

where $\beta_l = \|g_l - h_l\|_\infty$, $l \in \{1, \ldots, k\}$.
First by lemma 4, one can apply inequality (5.5) to the function g_k. We obtain:

$$\left|\int g_k d\mu - \int g_k d\nu\right| \leq \sum_{i \in I} (bD)_i \delta_i(g_k).$$

Using for $i \in I$, the inequality $\sum_{j \in I} C^n_{j,i} \leq c^n = L^n$ and the bound given in (5.3), we have:

$$(bD)_i \leq \sum_{j \in I} D_{j,i} \sup_{j \in I} b_j \leq \frac{b-a}{1-L} \sup_{t \in I} \sup_{(x,z) \in S \times \mathcal{X}} |\mu(p_t \leq x/z) - \nu(p_t \leq x/z)|, \tag{5.6}$$

Then using lemma 4, we conclude that:

$$\left|\int g_k d\mu - \int g_k d\nu\right| \leq M \frac{b-a}{1-L} \sup_{t \in I} \sup_{(x,z) \in S \times \mathcal{X}} |\mu(p_t \leq x/z) - \nu(p_t \leq x/z)|.$$

If we use the bound for β_k in Lemma 5, we conclude that:

$$\left|\int f_k d\mu - \int f_k d\nu\right| \leq C \sup_{t \in I} \sup_{(x,z) \in S \times \mathcal{X}} |\mu(p_t \leq x/z) - \nu(p_t \leq x/z)|,$$

with $C = 1 + M(b-a)\left(k - 1 + \frac{1}{1-L}\right)$. The proof of theorem 4 is now complete. $\square$

Lemma 3 *Let $g \in L(\mathcal{X})$ and $t \in I$. Then*

1. *$\mu_t(g) \in L(\mathcal{X})$ and:*
$$\sum_{i \in I} \delta_i(\mu_t(g)) \leq \sum_{i \in I} \delta_i(g).$$
2. $\|\mu_t(g) - \nu_t(g)\|_\infty \leq (b-a)\delta_t(g) \sup_{(x,z) \in S \times \mathcal{X}} |\mu(p_t \leq x/z) - \nu(p_t \leq x/z)|.$

Proof of lemma 3

1. For $(x, z, t) \in S \times \mathcal{X} \times \mathbb{Z}^2$, with $(xz)_t$ the element r of $\mathcal{X}$ such that $r_t = x$ and $r_s = z_s$ if $s \neq t$. Let $g_{z,t} : S \to \mathbb{R}$, $x \to g((xz)_t)$ is a Lipshitz function and then is derivable almost everywhere with $g'_{z,t}$ satisfying $\left\|g'_{z,t}\right\|_\infty \leq \delta_t(g)$. Let $z, \widetilde{z} \in \mathcal{X}$. With an integration by parts formula, we have:

$$\begin{aligned} A &= \left|\int g_{\widetilde{z},t}(x) \left(\mu(p_t \in dx/z) - \mu(p_t \in dx/\widetilde{z})\right)\right| \\ &= \left|\int \left(\mu(p_t < dx/z) - \mu(p_t < x/\widetilde{z})\right) g'_{z,t}(x) dx\right| \\ &\leq \delta_t(g) \sum_{i \neq t} L_{t,i} \left|z_i - \widetilde{z}_i\right|. \end{aligned}$$

Now this leads to:

$$
\begin{aligned}
& \left|\int g_{z,t}(x)\mu(p_t \in dx/z) - \int g_{\widetilde{z},t}(x)\mu(p_t \in dx/\widetilde{z})\right| \\
\leq \;& \sum_{i \neq t} \delta_i(g)\,|z_i - \widetilde{z}_i| + A \\
\leq \;& \sum_{i \neq t} \delta_i(g)\,|z_i - \widetilde{z}_i| + \delta_t(g) \sum_{i \neq t} L_{t,i}\,|z_i - \widetilde{z}_i|
\end{aligned}
$$

From this bound, it is obvious that $\mu_t(g) \in L(\mathcal{X})$ if $g \in L(\mathcal{X})$ since:

$$
\sum_{i \in I} \delta_i(\mu_t(g)) \leq \sum_{i \neq t} \delta_i(g) + L\delta_t(g) < \infty.
$$

2. For $z \in \mathcal{X}$, with another integration by parts formula:

$$
\begin{aligned}
|\mu_t(g)(z) - \nu_t(g)(z)| &= \left|\int g_{z,t}(x)\mu(p_t \in dx/z) - \int g_{z,t}(x)\nu(p_t \in dx/z)\right| \\
&\leq \left|\int \left(\mu(p_t < x/z) - \nu(p_t < x/z)\right) g'_{z,t}(x)dx\right| \\
&\leq (b-a)\delta_t(g) \sup_{(x,z) \in S \times \mathcal{X}} |\mu(p_t \leq x/z) - \nu(p_t \leq x/z)| \square
\end{aligned}
$$

Lemma 4 *For $l \in \mathbb{N}^*$, we have $g_l \in L(\mathcal{X})$ and:*

$$
\sum_{i \in I} \delta_i(g_l) \leq M.
$$

Proof of Lemma 4 First using assumption **(A5)**, $g_1 \in L(\mathcal{X})$ and we have:

$$
\sum_{i \in i} \delta_i(g_1) \leq \sum_{i \neq t_1} M_{t_1,i} \leq M.
$$

Using the point 1 of lemma 3, a straightforward finite induction shows that $g_l \in L(\mathcal{X})$, $l \leq k$. Now, if $l \geq 2$, with the point 1 in lemma 3 and the definition of g_l,

$$
\sum_{i \in I} \delta_i(g_l) \leq \sum_{i \in I} \delta_i(g_{l-1}),
$$

and Lemma 4 follows.$\square$

Lemma 5 *For $l \in \mathbb{N}^*$, we have:*

$$
\beta_k \leq (1 + (k-1)(b-a)M) \times \max_{t \in \{t_1,\dots,t_k\}} \sup_{(x,z) \in S \times \mathcal{X}} |\mu(p_t \leq x/z) - \nu(p_t \leq x/z)|.
$$

Proof of Lemma 5 We have by definition $\beta_1 \leq \sup_{(x,z) \in S \times \mathcal{X}} |\mu(p_{t_1} \leq x/z) - \nu(p_{t_1} \leq x/z)|$. Now, let $l \in \{1, \dots, k-1\}$. We have using the point 2 of Lemma 3 and Lemma 4:

$$
\begin{aligned}
\beta_{l+1} &= \left\|\mu_{t_{l+1}}(g_l) - \nu_{t_{l+1}}(h_l)\right\|_\infty \\
&\leq \left\|\nu_{t_{l+1}}(g_l) - \nu_{t_{l+1}}(h_l)\right\|_\infty + \left\|\nu_{t_{l+1}}(g_l) - \mu_{t_{l+1}}(g_l)\right\|_\infty \\
&\leq \beta_l + (b-a)\delta_{t_{l+1}}(g_l) \times \sup_{(x,z) \in S \times \mathcal{X}} \left|\mu(p_{t_{l+1}} \leq x/z) - \nu(p_{t_{l+1}} \leq x/z)\right| \\
&\leq \beta_{T,l} + (b-a)M \sup_{(x,z) \in S \times \mathcal{X}} \left|\mu(p_{t_{l+1}} \leq x/z) - \nu(p_{t_{l+1}} \leq x/z)\right|
\end{aligned}
$$

One can easily deduce the result.$\square$

References

[1] Abend, K., Harley, T. J. and Kanal, L. N. (1965) *Classification of binary random patterns* IEEE Trans. Inform. Theory IT11 538544.

[2] Ashikhmin, M. (2001) *Synthesizing Natural Textures.* The proceedings of 2001 ACM Symposium on Interactive 3D Graphics, Research Triangle Park, NorthCarolina March 19-21, pp. 217-226.

[3] De Bonet, J.S (1997) *Multiresolution sampling procedure for analysis and synthesis of texture images.* In Proc. 24th Annual Conference on Computer Graphics and Interactive Techniques 361-368. ACM Press, New York.

[4] Besag, J. (1986) *On the Statistical Analysis of Dirty Pictures.* J. R. Statis. Soc., Series B, Vol. 48, No. 3, pp. 259-302.

[5] Bickel, P., Levina, E. (2006) *Texture synthesis and nonparametric resampling of random fields* The Annals of Statistics, Vol. 34, No 4, 1751-1773.

[6] Cross, G. C., Jain, A. K. (1983) *Markov random field texture models* IEEE Transactions on Pattern Analysis and Machine Intelligence, 5(1):25-39.

[7] Dehardt, J. (1971) *Generalizations of the Glivenko-Cantelli theorem.* Ann. Math. Statist., 2050-2055.

[8] Doukhan, P. (1994) *Mixing: Properties and Examples.* Lecture Notes in Statist. 85. Springer, New York. MR1312160

[9] Efros, A. A., Freeman, W. T. (2001) *Image quilting for texture synthesis and transfer.* In Proc. 28th Annual Conference on Computer Graphics and Interactive Techniques 341-346. ACM Press, New York.

[10] Efros, A. A., Leung, T. (1999) *Texture synthesis by non-parametric sampling.* In Proc. IEEE International Conference on Computer Vision 2 1033-1038. IEEE Computer Soc., Washington.

[11] Fllmer, H. (1988) *Random Fields and Diffusion Processes.* Ecole dEt de Probabilits de St. Flour XVI. Lecture Notes in Mathematics 1362, Springer, 101-203.

[12] Geman, S., Geman, D. (1984) *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images.* IEEE Trans. Pattern Anal. Mach. Intell, 6, 721-741, 1984.

[13] Georgii, H.-O. (1988) *Gibbs measures and phase transitions.* de Gruyter, Berlin.

[14] Guyon, X. (1995) *Random fields on a network.* Springer, Berlin.

[15] Heeger, D.J., Bergen, J. R. (1995) *Pyramid-based texture analysis/synthesis.* In Proc. 22nd Annual Conference on Computer Graphics and Interactive Techniques 229-238. ACM Press, New York.

[16] Liang, L., Liu, C., Xu, Y., Guo, B., Shum, H.-Y. (2001) *Real-time texture synthesis by patch-based sampling.* Technical Report MSR-TR-2001-40, Microsoft Research.

[17] Luettgen, M.R., Karl, W.C, Wilsky, A.S, Tenney, R.R. (1993) *Multiscale Representations of Markov Random fields.* IEEE transactions on signal processing, vol. 41, n 12, 3377-3396.

[18] Paget, R. (1999) *Nonparametric Markov random field models for natural texture images.* PhD Thesis, University of Queensland.

[19] Paget, R., Longstaff, I. D. (1998) *Texture synthesis via a noncausal nonparametric multiscale Markov random field* IEEE Transactions on Image Processing, vol. 7, n 6, pp. 925-931.

[20] Paparoditis, E., Politis, D. (2002) *The local bootstrap for Markov processes.* J. Statist. Plann. Inference, n 108, pp. 301-328.

[21] Popat, K., Picard, R. (1993) *Novel cluster-based probability model for texture synthesis, classification, and compression.* In Proc. SPIE Visual Communications and Image Processing 2094 756-768. SPIE, Bellingham, WA.

[22] Rajarshi, M. (1990) *Bootstrap in Markov sequences based on estimates of transition density.* Ann. Inst. Math. Statist. 42 253-268.

[23] Simoncelli, E., Portilla, J. (1998) *Texture characterization via joint statistics of wavelet coefficient magnitudes.* In Proc. Fifth IEEE International Conference on Image Processing 1, 62-66. IEEE Computer Soc., Washington.

[24] Wei, L.-Y., Levoy, M. (2000) *Fast texture synthesis using tree-structured vector quantization.* In Proc. 27th Annual Conference on Computer Graphics and Interactive Techniques 479-488. ACM Press, New York.

[25] Zhu, S. C., Wu, Y. N., Mumford, D. (1998) *Filters, random fields and maximum entropy (FRAME)- toward a unified theory for texture modeling.* Internat. J. Computer Vision 27, 107-126.

9